The potential for industrial activity among EU regions

- an empirical analysis at the NUTS2 level

Martin Gornig and Axel Werwatz

FORLand-Working Paper 13 (2019)

Published by
DFG Research Unit 2569 FORLand, Humboldt-Universität zu Berlin
Unter den Linden 6, D-10099 Berlin
https://www.forland.hu-berlin.de

Tel +49 (30) 2093 46845, Email gabriele.wuerth@agrar.hu-berlin.de

Agricultural Land Markets – Efficiency and Regulation

The potential for industrial activity among EU regions

Martin Gornig* and Axel Werwatz**

Abstract

In the last decade, many parts of the world experienced severe increases in agricultural land prices. This price surge, however, did not take place evenly in space and time. To better understand the spatial and temporal behavior of land prices, we employ a price diffusion model that combines features of market integration models and spatial econometric models. An application of this model to farmland prices in Germany shows that prices on a county-level are cointegrated. Apart from convergence towards a long-run equilibrium, we find that price transmission also proceeds through short-term adjustments caused by neighboring regions.

Keywords: Agricultural land markets; price diffusion; spatial dependence; ripple effect

JEL codes: Q 24, C 23

Acknowledgements

We gratefully acknowledge funding of this research from Deutsche Forschungsgemeinschaft (DFG) via Research Unit 2569 “Agricultural Land Markets – Efficiency and Regulation” and the Hans-Böckler-Stiftung via research project "Industry in the City". We would also like to thank the Advisory Board of this project for valuable comments and suggestions.

* Martin Gornig, mgornig@diw.de
Deutsches Institut für Wirtschaftsforschung, Abteilung Unternehmen und Märkte
Mohrenstraße 58, 10117 Berlin

** Axel Werwatz, axel.werwatz@tu-berlin.de,
Technische Universität Berlin, Fakultät VII - Wirtschaft und Management, Institut für Volkswirtschaftslehre und Wirtschaftsrecht, FG Ökonometrie und Wirtschaftsstatistik
Sekr. H 57, Straße des 17. Juni 135, 10623 Berlin

1 Introduction

The increased importance of the service sector and the concurrent loss of industrial jobs has been one of the salient features of economic development in advanced countries. With the rise of the new industrial giants China and India, the demise of industry in the West has appeared as a logical and inescapable trend. The global financial crisis of 2007 and 2008, however, has rekindled the interest of observed and policy makers in industry in Europe and elsewhere. Indeed, in 2012 the European Commission has called for a „European Industrial Renaissance" and put forward the target that the share of manufacturing in GDP reaches 20% by 2020.

While this target has been formulated at the European level, industrial activity varies widely across the regions of the European Union. And while national and European framework policies influence, to some degree, all regions, it is the regional conditions and policies that play a crucial role for whether a "European Industrial Renaissance" will indeed happen.

This is the perspective taken up in this paper, which analyzes the development of industrial acitivity at the level of EU NUT2 regions in the period from 2000 to 2015. We estimate the share of industrial output that a region can be expected to have in 2015, given the `exogenous' condition it faces. This allows us to estimate and analyze the deviation of a region's actual industrial share from its expected ("fair") industrial share and identify over- and underperformers. Moreover, we do not only estimate the level of regional industrial performance at the end of our time window. We also study how regions got there by documenting and analyze the heterogeneity among regions with regard to starting levels and trends.

To estimate the expected share of industrial output for each region we use a regression model based on a logistic trend function. This trend function is able to capture the overall tendency of declining industrial activity with a few parameters. It is the building block of our nonlinear regression model that also accounts for the main determinants of industrial output that can be regarded as largely exogenous from the point of view of regional policy makers: national framework conditions, geographical position of the region and its population density.

The remainder of this paper is organized as follows. In the next section, we describe the theoretical framework for building a statistical model of the expected industrial share of a region. We then turn towards the methodology we apply for estimating expected industrial shares. After describing the data, we proceed to the description of our empirical results. The final section concludes.

2 Theoretical Framework

In order to estimate a reference value of the industry share for each region, we need to discuss the circumstances that determine the level of industrial activity that can reasonably be expected for a region – circumstances, that are largely exogenous from a region's perspective.

European framework conditions, at least to a first approximation, equally affect each region and can thus be absorbed into the "constant term" of our model of the expectation. National framework conditions, however, differ between regions from different countries. Indeed, the new institutional economics, stresses the importance of the complex interaction of the many institutions shaping advanced societies for economic development at all levels. Because we neither aim to identify the separate effect of each institution on the regional industrial share nor are we able to empirically represent each of them, we will capture the effect of the set of national institutions simply by including country level dummy variables.

A theoretical perspective on regional (industrial) development with an explicit geographical focus is offered by the New Economic Geography (Krugman 1991). It stresses, among other things, the importance of access to markets and distance-related transaction costs (Krugman/Venables 1995; Fujita et al. 1999) for a region's ability to specialize and strive in industrial activity. The core zone of the European common market, which still absorbs a large share of EU industrial output, is the `Blue Banana' (RECLUS 1989), an eponymously shaped central area of densely populated and highly urbanized regions comprising London, Amsterdam, Brussels, Frankfurt and Milan. Being located within or close to this core area should offer the favorable circumstances for economic activity. Below, we thus include measures of the distance to the Blue Banana in our regression model of the expected industrial share of regions.

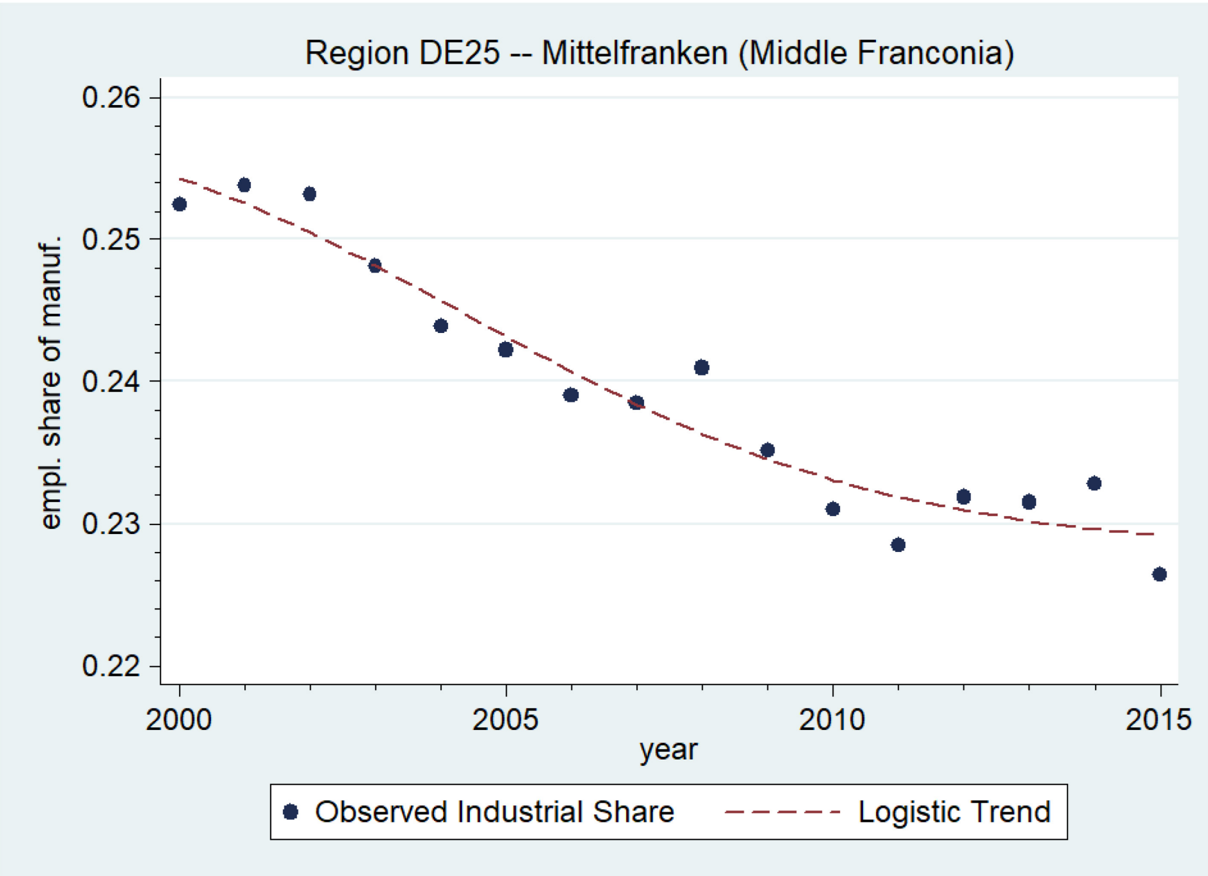
That densely populated areas enjoy agglomeration advantages is a key proposition of agglomeration theories. Firms from all sectors benefit from the highly developed infrastructures that typically can be found in large agglomerations. It is particularly the knowledge intensive sectors that profit from the local accumulation of education and research institutions and a highly skilled workforce. This will include most manufacturing sectors in the advanced European economies. However, industrial activity is space-consuming and thus necessarily competes with the many other valuable uses of the scarce territory of agglomerations. The relationship between the population density of a region and its share of industrial activity is thus less clear, even if agglomerations offer advantages especially for knowledge-intensive industrial firms. We still include measures of agglomeration or density among the exogenous variables in our regression model, as they are not easily altered, at least in the short- to medium run.

3 Method

The key to our analysis of the industrial activity among region is an estimate of the industrial share a region can be expected to have in our terminal year 2015, given the `exogenous' conditions outlined in the previous section. The regression model we use to obtain an estimate of this expected industrial share (given national framework conditions, geographical location and population density) is described in this section. Because our aim is not to accurately predict the industrial share of a region in 2015 but rather to derive a reasonable estimate of its expected share, given exogenous framework conditions, we aim at a parsimonious, parametric model.

The parameters of this model will be estimated from a data set that has information on each region's industrial share from 2000 to 2015. As an example, Figure 1 shows the development of the dependent variable, the industrial share, for the nuts2 region DE25 ("Mittelfranken" Middle Franconia). The data points clearly exhibit the general trend of a declining industrial share in advanced economies at the level of an individual region.

Figure 1: Development of the industrial share in region DE25 (Mittelfranken, Middle Franconia)



This declining trend is captured in the graph by the dashed line. It is the graph of a logistic trend function fit to the data for this region.

This logistic trend function is the starting point of our regression model and has the following general formula:

$$\mathrm{Trend}_t = b_0 + b_1 / \{1 + \exp[-b_2(t - b_3)]\}$$

where the time index t denotes a year in our context. The meaning of the parameters b_0, b_1, b_2 and b_3 is illustrated in Figure 2.

Figure 2: Illustration of the meaning of the parameters of the logistic trend function.

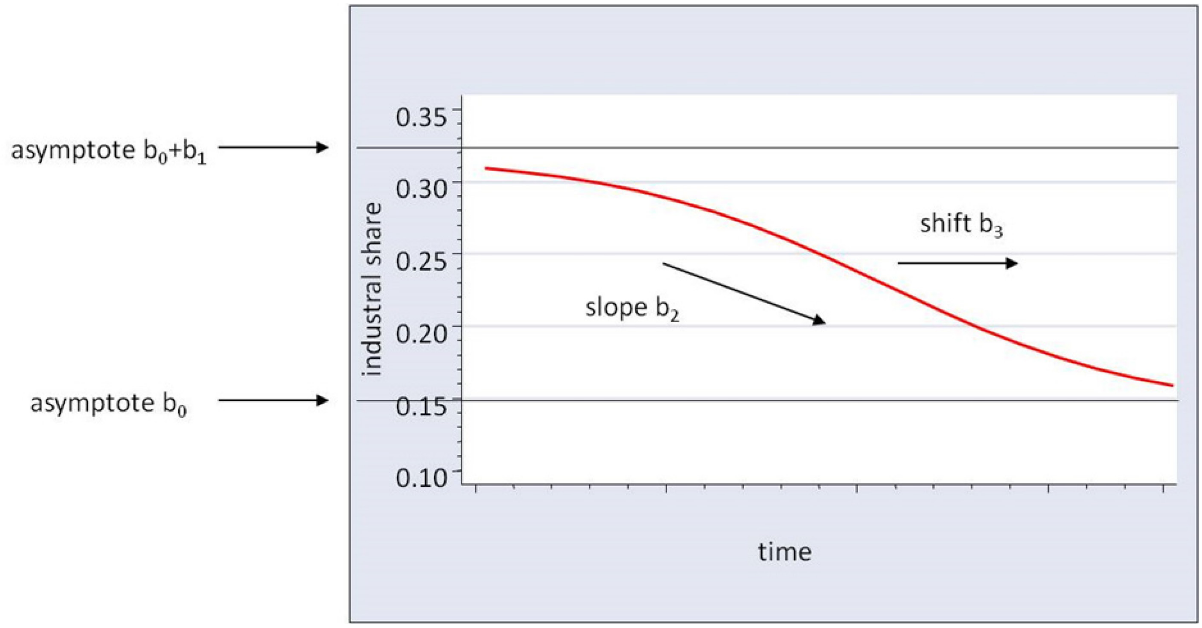


Figure 2 shows the ability of the logistic trend function to capture a declining trend (via the slope parameter b_2) that may eventually level out and approach a lower asymptote (represented by the parameter b_0). This is precisely the pattern observed for the industrial share of the Mittelfranken region in Figure 1 and that can indeed be observed for most other European regions. The logistic trend function also allows for an upper asymptote (given by the sum of b_0 and b_1) as well as a shift along the time axis (via the parameter b_3).

With just four parameters the logistic trend function thus offers a parsimonious yet quite flexible way to capture the overall trend in our dependent variable in the 2000 to 2015 period. It is thus the first part on the right hand side of our regression model of industrial shares given by the following equation

$$Y_{jt} = \underbrace{\underbrace{b_0 + \frac{b_1}{1+\exp[-b_2(t-b_3)]}}_{\text{trend}} \quad + \quad \underbrace{\alpha \cdot \text{Location}_{jt} + \beta \cdot \text{Density}_{jt} + \gamma_c \cdot \text{CountryC}_j}_{\text{exogenous determinants}}}_{\text{expected industry share}} + \underbrace{\varepsilon_{jt}}_{\text{deviation}} \tag{1}$$

Here, Y_{jt} denotes the industrial share of region j at time t. The right hand side of regression equation (1) starts with the logistic trend and adds terms for the three exogenous determinants. The precise definition of the location and density variables will be given below in the data section. The effect of the national framework conditions is picked up by dummy variables for each country. In equation (1) it is assumed that region j belongs to country “C” and that the parameter γ_c measures the effect of country C on the industrial share of its regions (relative to the left-out reference country).

4 Estimation

The simple specification in (1), where location and density enter linearly, is made more flexible in the empirical analysis as explained below. Before we turn to such model specification and model selection issues, we want to describe our estimation method using the simple specification of equation (1). We will estimate the parameters of this (and all other variants of the model) with the nonlinear least squares (NLSQ) method. That is, we minimize the following criterion function with respect to the regression parameters b_0, b_1, b_2, b_3, α, β, γ_N, γ_S, and γ_E is:

$$Min \sum_t \sum_j \left[Y_{jt} - \left(b_0 + \frac{b_1}{1+\exp[-b_2(t-b_3)]} + \alpha \cdot \text{Location}_{jt} + \beta \cdot \text{Density}_{jt} + \gamma_N \cdot \text{N}_{jt} + \gamma_S \cdot \text{S}_{jt} + \gamma_E \cdot \text{E}_{jt} \right) \right]^2$$

For simplicity, we have assumed here that there are only four countries, named „N“, „S“, „E“ and „W“ and that country „W“ is the reference category. The parameters that minimize the squared distances of the observed industry shares from their expected values according to our nonlinear model are found iteratively.

5 Model Selection

The baseline specification (1) includes the continuous regressors location and distance in the simplest possible way: as additive, linear predictors whose effect is given by their coefficients. The effect of these variables, however, may well change with their level. We allow for the possibility of such a nonlinear influence in two ways: parametric polynomials of higher order and step-functions (see Figure 14 below).

The different alternatives for modeling the effect of location and distance can be combined with each other in multiple ways. A comprehensive list of the possible alternatives for the right hand side of our regression model is given in Table 4 below. For selecting the best specification among these alternatives we use adjusted R^2. We do not use a cross-validation criterion because we do not aim at prediction and because we do not have the necessary sample size for resampling based model selection. Adjusted R^2 is easily calculated and reasonably captures adjusts in-sample fit for model size.

6 Data

Our analysis of the industrial share of European regions is conducted at the level of the 262 NUTS2 regions, using annual data for each region between 2000 and 2015. In this section, we show that NUTS2 regions vary considerably regarding their industrial share, density and location and what alternative ways for measuring these quantities are considered in our empirical analysis.

6.1 Dependent Variable

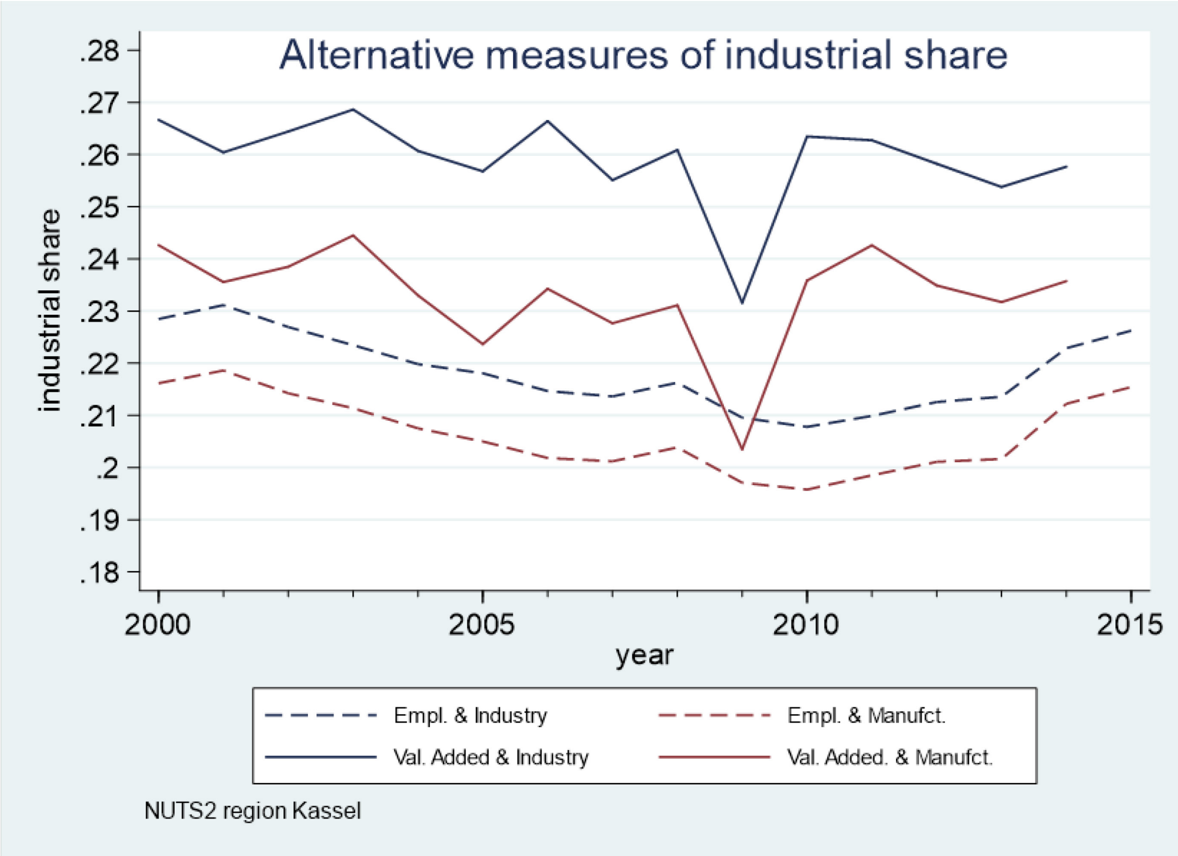
There are several issues regarding the measurement of the industrial share of a region. The first issue is the delineation of the industrial sector for statistical purposes. Industry more narrowly defined is “manufacturing”. Manufacturing, according to EUROSTAT, “includes the physical or chemical transformation of materials, substances, or components into new products”. Statistically, it is comprised of all activities in Section C of the EU NACE classification (Rev.2). A wider definition also contains Mining and Quarrying (Section B) and Utilities (Section D). We will refer to the latter as „industry“. The second issue is whether industrial activity should be measured in terms of output (value added) or employment.

The different possibilities are depicted in Figure 3, using the German region “Kassel” for illustration purposes. Blue lines are based on the wider definition of the industrial sector, red lines are for the narrower „manufacturing“ definition. Solid lines use employment relative to population as the industrial share whereas the dashed lines show the industrial share of value added in the Kassel region. Two things clearly emerge from Figure 3:

(1) as is to be expected, the wider definition of the industrial sector leads to slightly higher industrial shares and

(2) the industrial shares based on value added are considerably more volatile.

Figure 3: Development of the industrial share in the region Kassel (DE73) according to four different definitions.



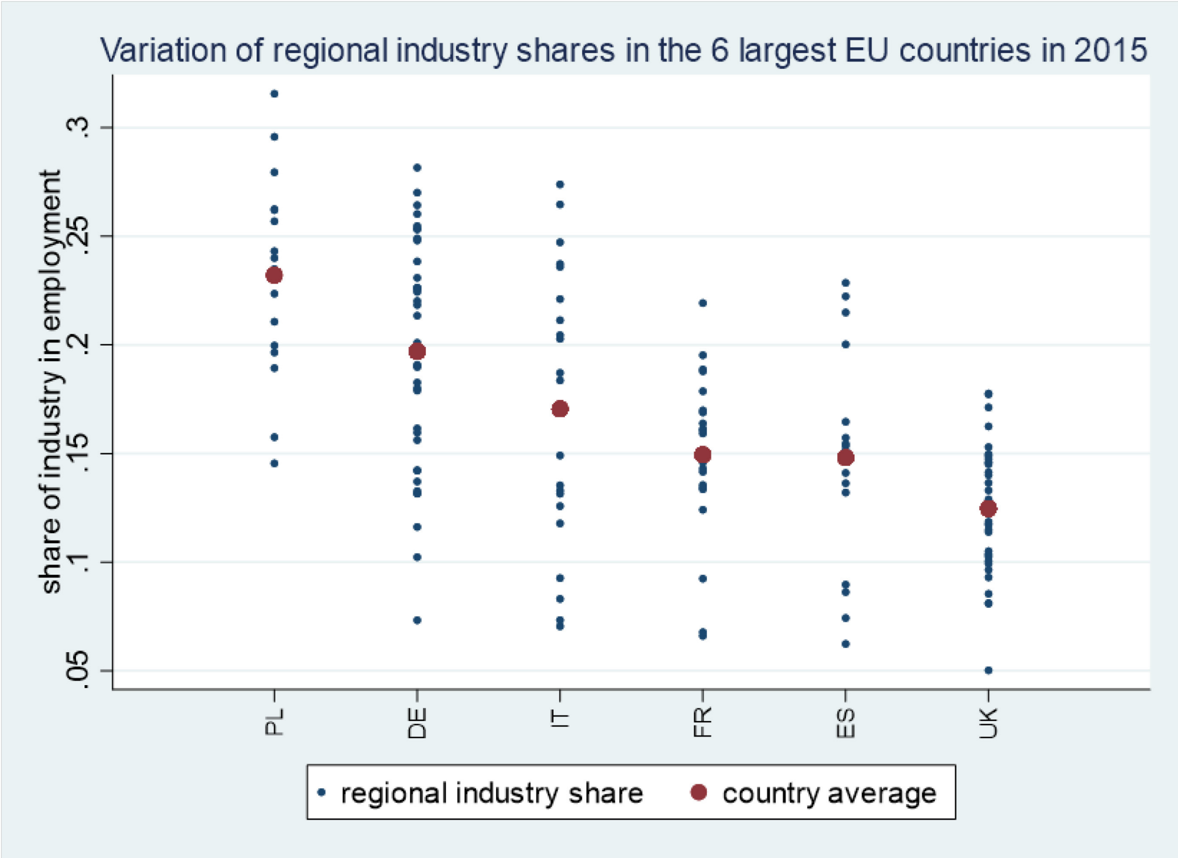
Because transitory random fluctuations are more present in the value-added based measures of industrial activity we will be working with the employment based shares in the empirical analysis. Moreover, our focus will be on the wider definition of industry.

6.2 Exogenous variables

In order to set the stage for our exogenous variables, we show a portion of the variation of our dependent variable in Figure 4. While Figure 1 and Figure 3 showed the variation in the industrial share for particular regions over time, Figure 4 exhibits both the variation of the industrial shares within and between the six largest EU countries at a given point in time (2015). For each country, the blue dots are stacked up in a pillar and show the variation of the industrial share among the regions of that country in 2015. We also included, as a reference point, a large red dot representing a country's average industrial share in 2015. We used the value of this average to order the countries by decreasing average industrial shares with Poland at the top left end and the UK at the bottow right end of the spectrum.

Apparently, Poland's average industrial share (≈23%) in 2015 is considerably higher than that of the UK (≈12%). However, while Polish regions tend to have higher industrial shares, there is considerable variation among them. Hence, it is not only national framework conditions that matter for a region's industrial share but also region specific determinants, such as location and density. The distribution of these determinants in our data is considered next.

Figure 4: Regional industry shares in the six largest EU countries in 2015.



6.3 National framework conditions

As already mentioned above, national framework conditions will simply be represented by country-level dummy variables in our model. We will denote these dummy variables with the country codes as illustrated in the first column of Table 1. In this table we give the sample averages of the country dummy variables for the six largest EU countries, which –of course– simply amount to the fraction of all EU regions stemming from the particular country.

Table 1: Share of each of the six largest EU-countries among the NUTS2 regions

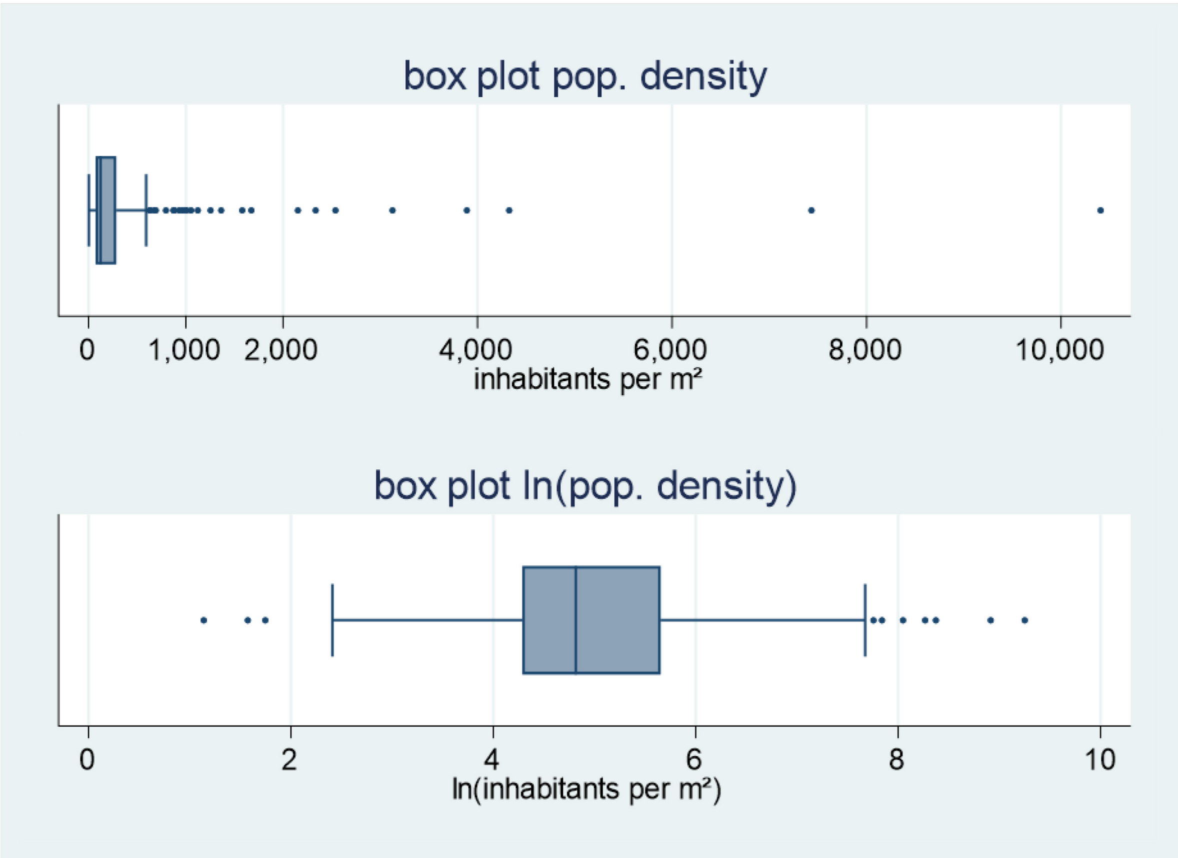
Variable	Anteil	Anzahl
DE	0.145	38
UK	0.137	36
FR	0.084	22
IT	0.080	21
ES	0.061	16
PL	0.061	16

We find that 14.5% of all NUTS2 regions are located in Germany (38 of 262 total regions). The UK comes in second accounting for 13.7% of all regions.

6.4 Population Density

Agglomerations are particularly attractive for knowledge-intensive firms but offer considerable benefits for all sorts of economic activity via the size of the local market or the well-developed infrastructure. We do not aim at measuring these benefits separately and quantifying their separate effects on the industrial share. Rather, we want to employ an overall measure for agglomerative advantages. One such measure is population density. The distribution of population density is depicted by the box plot in the upper panel of Figure 5. Apparently, the distribution is strongly skewed with a long right tail. The central 50% of the regions, that are contained in the blue shaded box have a population density ranging from about 70 inhabitants per m² to 280 inhabitants per m². There are, however, also some regions with densities beyond 1000 inhabitants per m² like Ile de France (1005 E/m²), Berlin (3890 E/m²), Région de Bruxelles-Capitale (7433 E/m²) and Greater London (10406 E/m²).

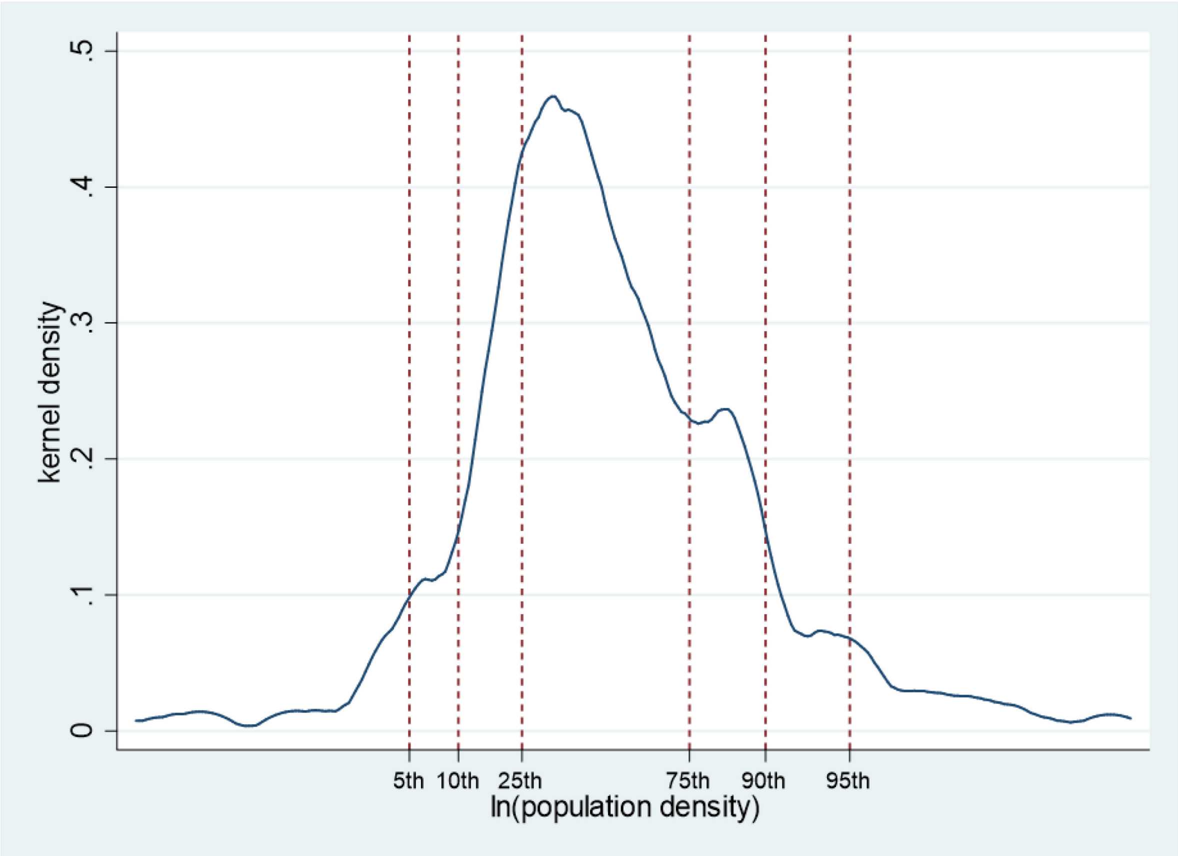
Figure 5: Box plots of the distribution of the population density (upper panel) and the log population density (lower panel) among the NUTS2 regions.



The lower panel of Figure 5 shows that the distribution becomes much more symmetric once we apply the log-transformation to the population density. This is why we will be using log-population density in the empirical analysis as a continuous regressor.

However, for both the population density and the log population density we create and employ discrete versions as alternatives. This is illustrated in Figure 6, which shows a kernel density estimate of the distribution of log population density in 2015. The vertical dashed lines indicate the 5th, 10th, 25th, 75th, 90th and 95th percentiles of the distribution, respectively. These can be used to define seven categories, corresponding to the intervals formed by these percentiles. For each category, a dummy variable can be defined, indicating whether a region's log population density belongs, say, to the lowest 5 percent of the distribution (category 1). These dummy variables can be included as a regressors as an alternative to the continuous log population density variable. We similarly subdivide and create categories for the (unlogged) population density.

Figure 6: Kernel density estimate of the distribution of log population density among the NUTS2 regions. Dashed vertical lines mark the indicated percentiles.



6.5 Location

Trade theory stresses the importance of location for the export potential of a region and thus its expected potential for industrial activity. We measure location by the distance of a region to the „Blue Banana“, the core area of the European common market shown in Figure 7. This distance is calculated as the distance between the largest city of a region and the closest large city within the „Blue Banana“. For the NUTS2 Region DE40 (Brandenburg), for instance, this is the distance between the cities of Potsdam (Brandenburg’s capital) and Kassel (the closest large city within the blue banana), which amounts to 350 km.

Figure 7: Location of the central EU area referred to as the "Blue Banana".

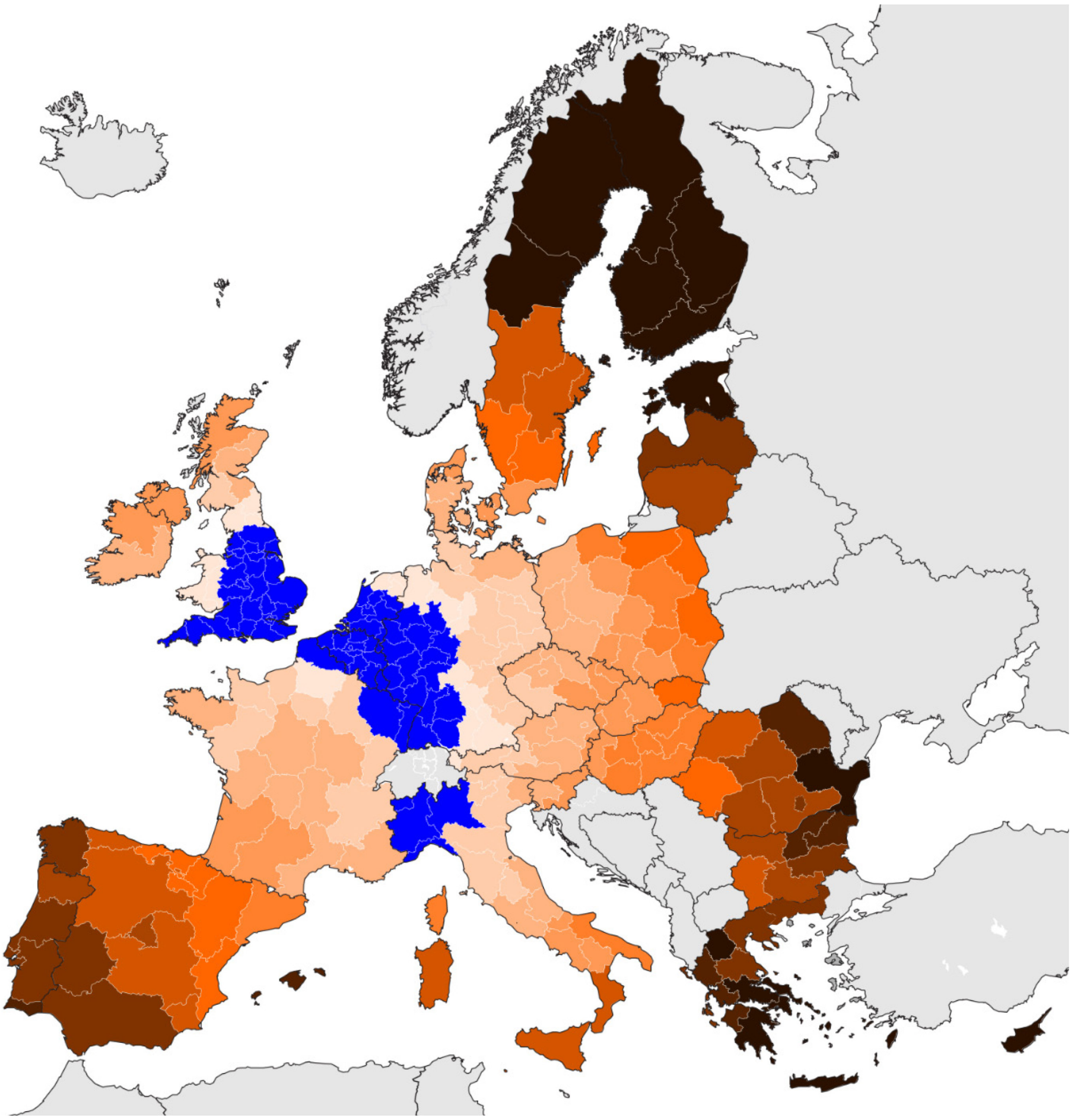


Figure 8 shows the distribution of the distance to the „Blue Banana“ as a kernel density. The median distance of this right-skewed distribution is at approximately 500 km (vertical blue finely dashed line). We employ this variable in the empirical analysis as a continuous exogenous variable and alternatively in categorical form. The categories are formed according to the intervals represented by the red vertical dashed lines in the graph. We defined rather narrow intervals in the dense left part of the distribution. Details of the definitions of the categories are given in Table 2:

Figure 8: Kernel density estimate of the distribution of the distance to the "Blue Banana".

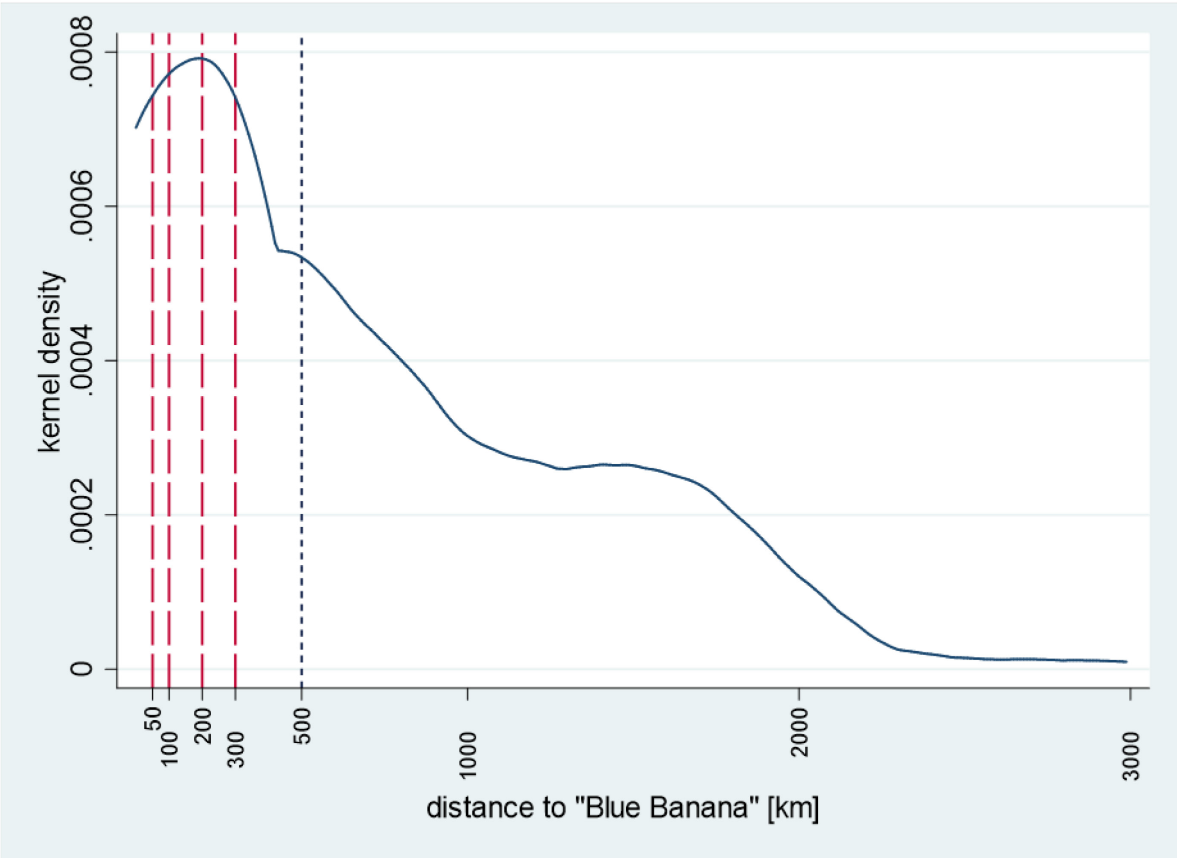


Table 2: Definition of distance categories

Distance categories	Definition
dist_type 1	less than 50 km distance to Blue Banana
dist_type 2	50 to 100 km
dist_type 3	100 to 200 km
dist_type 4	200 to 300 km
dist-type 5	more than 300 km distance

Instead of using the geographical distance to the „Blue Banana“, location advantages of a region could be measured in terms of the travel time to the „Blue Banana.“ In the example already used above for illustration purposes, this amounts to a „time distance“ between Potsdam and Kassel of 3 hours and 30 minutes. This travel time puts the region DE40

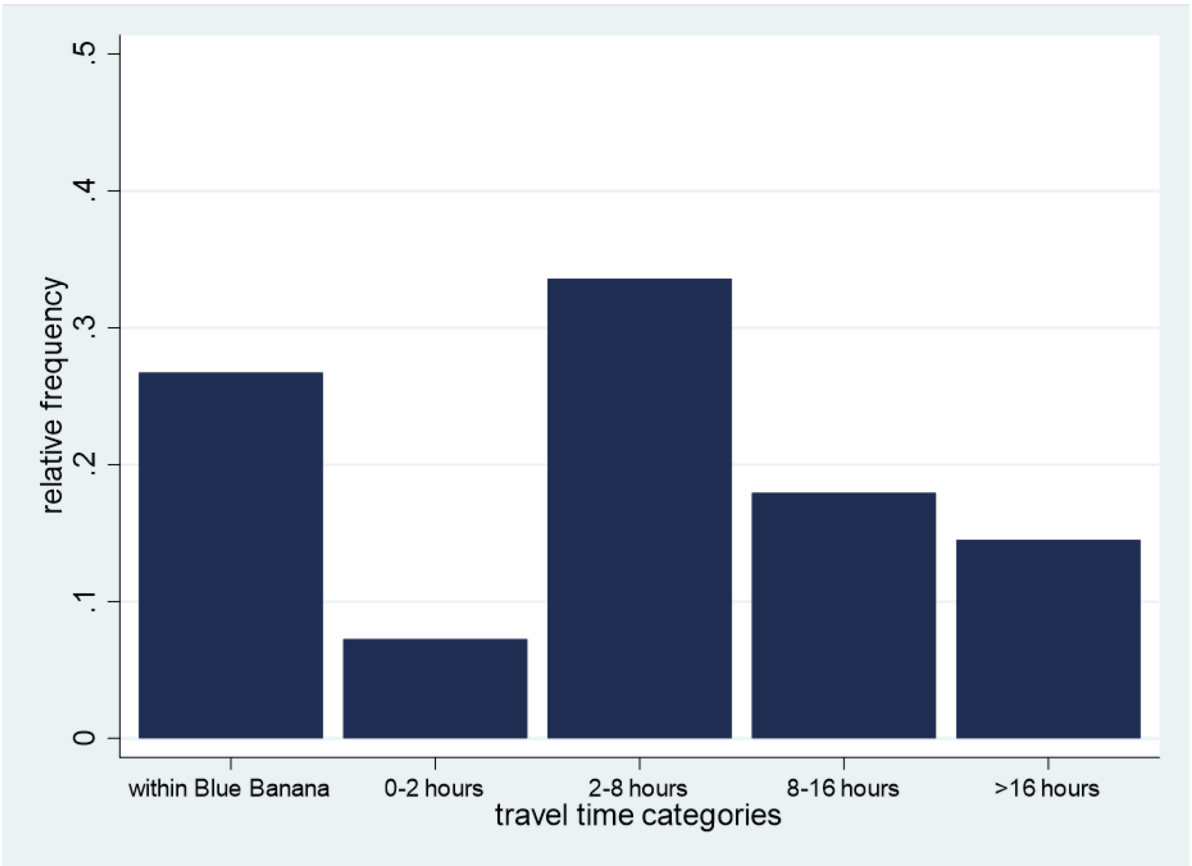
(Brandenburg) into the 3rd category of the following classification of “access groups” (see Table 3), the regions with a travel time to the “Blue Banana” between 2 and 8 hours.

Table 3: Definition of travel time intervals and corresponding „access“ dummy variables

Travel time to Blue Banana	Dummy Variable
inside Blue Banana	access_0
0 – 2 hours	access_0_2
2 – 8 hours	access_2_8
8 – 16 hours	access_8_16
more than 16 hours	access_16

For regions in this group, the dummy variable access_2_8 will take on the value 1 and will equal 0 for all other regions. The other dummy variables are similarly defined and named, as can be seen from the second column of Table 3. Figure 9 shows the resulting frequency distribution of the access (to the „Blue Banana“) categories.

Figure 9: Frequency distribution of the five „access“ categories.



Most regions belong to the 3rd category, with a travel time to the „Blue Banana“ between 2 and 8 hours. Regions within the „Blue Banana“ (access_0=1) account for roughly a quarter of all NUTS2 regions.

7 Empirical Results

The aim of our empirical analysis is to estimate the industrial share of a region in 2015, given its exogenous conditions. The regression model we employ for this purpose is deliberately parsimonious. It is not meant to deliver accurate predictions of a regions actual industrial share at the end our observation window. Instead, it is based on a theoretical perspective: which industrial share can reasonably be expected from a region given the three main determinants that –as a first approximation- are beyond its influence: national framework conditions, location and population density. Our baseline model, which is repeated here for convenience, is a particularly simple way of obtaining such an expected industrial share:

$$Y_{jt} = \underbrace{\underbrace{b_0 + \frac{b_1}{1+\exp[-b_2(t-b_3)]}}_{\text{trend}} + \underbrace{\alpha \cdot \text{location}_{jt} + \beta \cdot \text{density}_{jt} + \gamma_c \cdot \text{countryC}_j}_{\text{exogenous determinants}}}_{\text{expected industry share}} + \underbrace{\varepsilon_{jt}}_{\text{deviation}}$$

Given estimates of the parameters on the right hand side, an estimated expected industry share can be obtained from

$$\hat{E}[Y_{\text{AT11, 2015}} \mid \text{loc} = 644, \text{den} = 73] = \hat{b}_0 + \frac{\hat{b}_1}{1+\exp\left[-\hat{b}_2\left(2015-\hat{b}_3\right)\right]} + \hat{\alpha} \cdot 644 + \hat{\beta} \cdot 74 + \hat{\gamma}_{AT} \qquad (2)$$

where we have used Austria's Burgenland region for illustration purposes, which has a distance from the "Blue Banana" of 644 km and a population density of about 73 inhabitants per m^2. All coefficients on the right hand side of equation (2) have a "hat" superscript to indicate that they represent estimates from sample data in this formula. Similarly, on the left hand side, we have put a hat on the expectation symbol *E* to indicate that equation (2) would yield an estimate of the expected industry share of this region in 2015.

Once we have obtained such an estimate for a region, we can proceed to calculate the deviation of its observed industrial share in 2015 from its estimated expected ("potential") industry share. Formally,

$$\hat{\varepsilon}_{\text{AT11, 2015}} = Y_{\text{AT11, 2015}} - \hat{E}[Y_{\text{AT11, 2015}} \mid \text{loc} = 644, \text{den} = 73] \qquad (3)$$

This „residual" can be regarded as an estimate of the deviation ε_{jt} in equation (1) for the particular region at the indicated time period. It is an estimate of how much it exceeds ($\hat{\varepsilon}_{jt}$>0) or falls short ($\hat{\varepsilon}_{jt}$<0) of its expected share, given its exogenous circumstances. Regions with large positive values of $\hat{\varepsilon}_{jt}$ can then be regarded as "over achievers" or "high performers" whereas regions with large negative values of $\hat{\varepsilon}_{jt}$ can then be regarded as "under achievers" or "low performers".

7.1 Model Selection

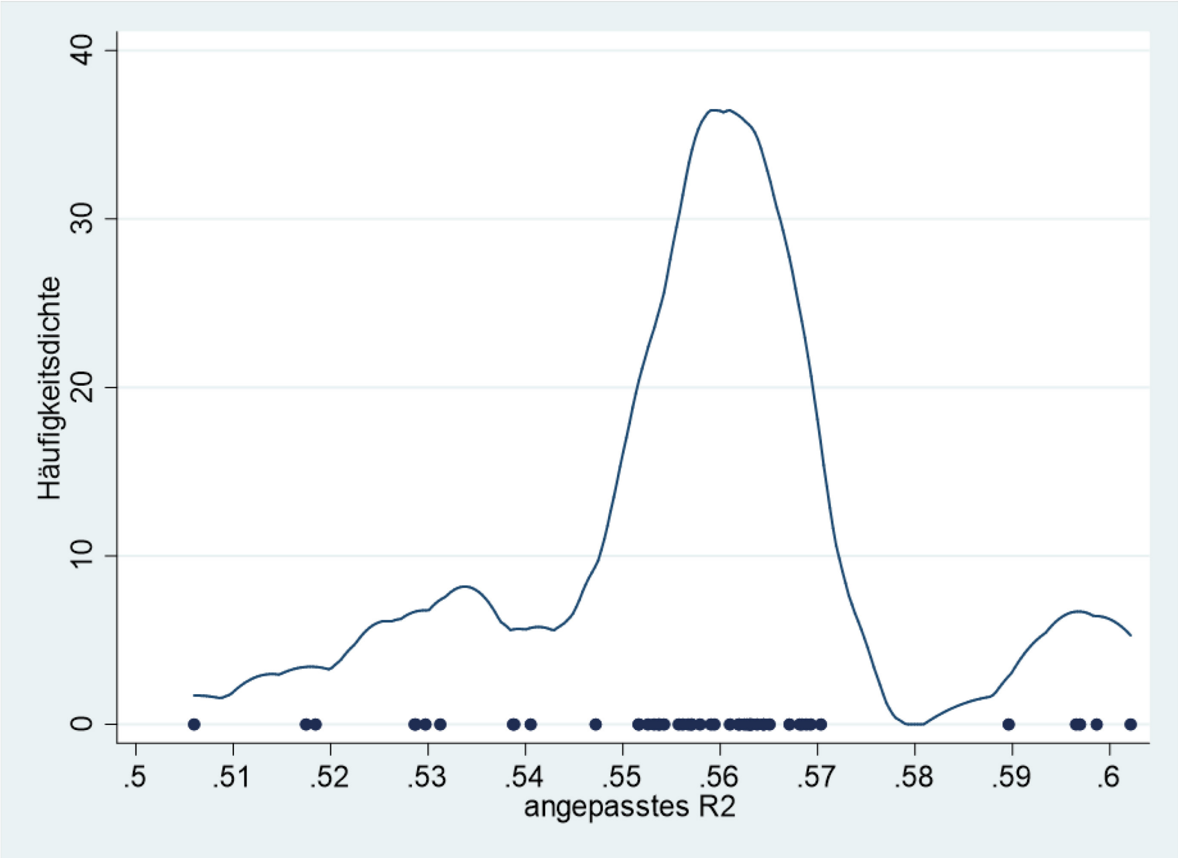
Our regression deliberately does not contain many “predictors” on the right hand side. We therefore do not have the variable selection problem that is a crucial component of using regression for accurate prediction. As already indicated in the Methods section, we still have various alternative measures for location and population density that can enter the model in continuous or discrete form. An overview of the modeling and measurement alternatives we considered for the effect of location and population density is given in the following Table 4:

Table 4: Alternatives considered for modeling the effect s of location and population density

Location	**Density**
Linear effect of distance $\alpha \cdot \text{distance}_{jt}$	Linear effect of log population density $\beta \cdot \ln(\text{density}_{jt})$
Quadratic effect of distance $\alpha_1 \cdot \text{distance}_{jt} + \alpha_2 \cdot \text{distance}_{jjt}^2$	Quadratic effect of log population density $\beta_1 \cdot \ln(\text{density}_{jt}) + \beta_2 \cdot \ln(\text{density}_{jt})^2$
Cubic effect of distance $\alpha_1 \cdot \text{distance}_{jt} + \alpha_2 \cdot \text{distance}_{jt}^2 + \alpha_3 \cdot \text{distance}_{jt}^3$	Cubic effect of log population density $\beta_1 \cdot \ln(\text{density}_{jt}) + \beta_2 \cdot \ln(\text{density}_{jt})^2 + \beta_3 \cdot \ln(\text{density}_{jt})^3$
Quartic effect of distance $\alpha_1 \cdot \text{distance}_{jt} + \alpha_2 \cdot \text{distance}_{jt}^2 + \alpha_3 \cdot \text{distance}_{jt}^3 + \alpha_4 \cdot \text{distance}_{jt}^4$	Quartic effect of log population density $\beta_1 \ln(\text{density}_{jt}) + \beta_2 \ln(\text{density}_{jt})^2 + \beta_3 \ln(\text{density}_{jt})^3 + \beta_4 \ln(\text{density}_{jt})^4$
Linear effect of log distance $\alpha \cdot \ln(\text{distance}_{jt})$	log population density group dummies $\beta_1 \text{ln\_dens\_gr2} + \beta_2 \text{ln\_dens\_gr3} + \beta_3 \text{ln\_dens\_gr4} + \beta_4 \text{ln\_dens\_gr5}$
Quadratic effect of log distance $\alpha_1 \cdot \ln(\text{distance}_{jt}) + \alpha_2 \cdot \ln(\text{distance}_{jt})^2$	population density group dummies $\beta_1 \text{dens\_gr2} + \beta_2 \text{dens\_gr3} + \beta_3 \text{dens\_gr4} + \beta_4 \text{dens\_gr5}$
Cubic effect of log distance $\alpha_1 \cdot \ln(\text{distance}_{jt}) + \alpha_2 \cdot \ln(\text{distance}_{jt})^2 + \alpha_3 \cdot \ln(\text{distance}_{jt})^3$	
Quartic effect of log distance $\alpha_1 \ln(\text{distance}_{jt}) + \alpha_2 \ln(\text{distance}_{jt})^2 + \alpha_3 \ln(\text{distance}_{jt})^3 + \alpha_4 \ln(\text{distance}_{jt})^4$	
Distance group dummies $\alpha_1 \cdot \text{dist\_type2} + \alpha_2 \cdot \text{dist\_type3} + \alpha_3 \cdot \text{dist\_type4} + \alpha_4 \cdot \text{dist\_type5}$	
Time to „Blue Banana“ access dummies $\alpha_1 \cdot \text{access\_0\_2} + \alpha_2 \cdot \text{access\_2\_8} + \alpha_3 \cdot \text{access\_8\_16} + \alpha_4 \cdot \text{access\_16}$	

The ten modeling alternatives for the effect of location (=distance) in the left column of Table 4 can be combined with each of the five modeling alternatives for the effect of population density in the right column, leading to a total of 10 x 5 = 50 alternative specifications of the right hand side of our regression model. We estimated each of these 50 specifications with the method of nonlinear least squares as described above and calculated the corresponding adjusted R^2. The resulting distribution of adjusted R^2 is depicted in Figure 10, which also shows the individual realizations as blue dots on the horizontal axis.

Figure 10: Distribution of the adjusted R2 value among the 50 alternative specifications.



Most specifications achieve an adjusted R^2 around 56% but a few manage to even attain adjusted R^2 values around 60%. The specifications behind the cluster at the upper right end of Figure 10 all model the influence of location via the dummy variables for time-to-Blue-Banana intervals. Regarding the influence of population density, no clear cut „winner“ specification emerges. The specification with a quartic polynomial of log population density does about as well as the variant which discretizes this variable into dummy variables. We therefore present the results of each of these specifications in Table 5.

7.2 Estimated expected industry shares

Table 5 includes, in addition to the estimated coefficients, standard errors and t-ratios. The latter generally exceed 2 in absolute value for both specifications, indicating that virtually all coefficients tend to be statistically significant at conventional levels. Moreover, the size and direction of the estimated coefficients of the two specifications are very similar to each other. Only the estimated b0 coefficient of the logistic time trend is considerably larger for model 1 than for model 2. This stems from the fact that b0 represents the (hypothetical) industy share of a region where all exogenous variables equal 0. In model 1 this would be a region with a log population density of 0 or a conventional population density of 1. Such a sparsely populated region cannot be found in the data. Since the industrial share is negatively related to log population density (see also below), b_0 is „artificially“ large in model 1.

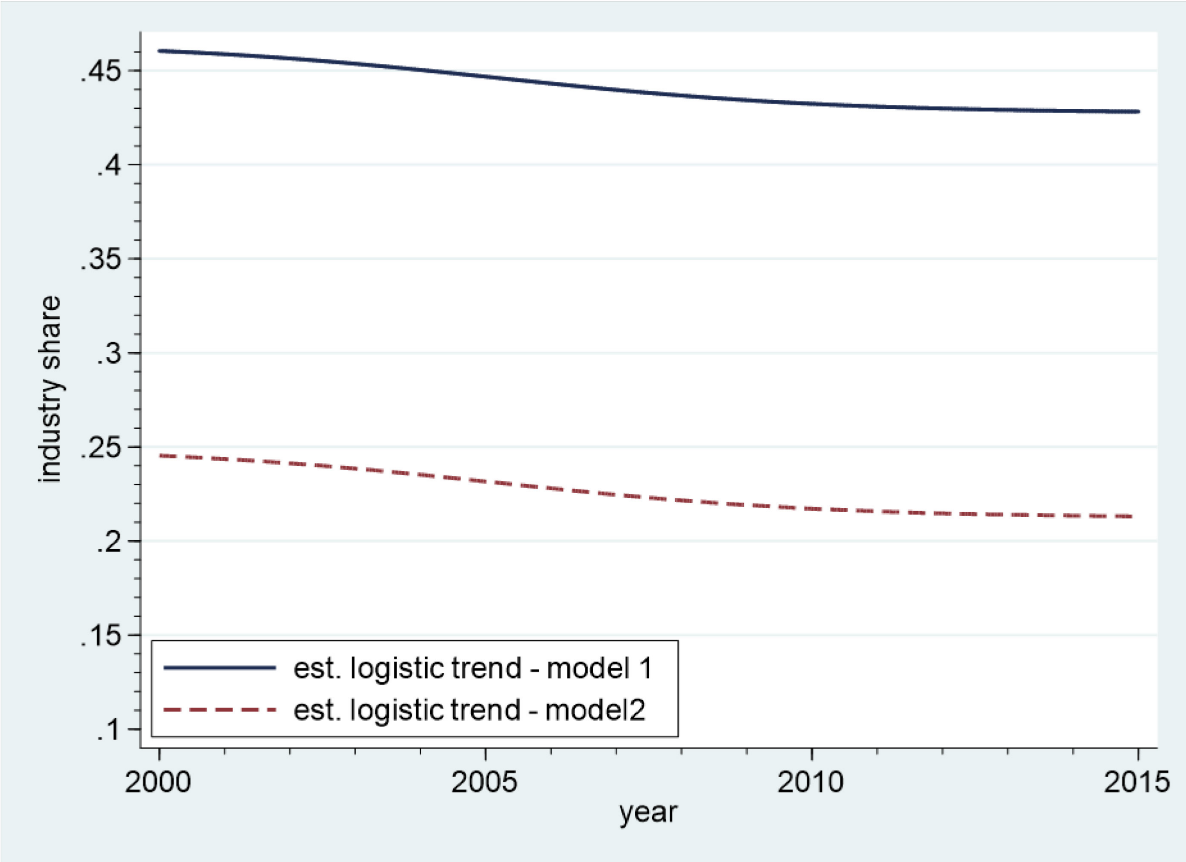
Table 5: Nonlinear least squares estimates of two specifications with high adjusted R^2 values.

	model 1			model 2			
variable	**coef.**	**std. err.**	**t**	**coef.**	**std.err.**	**t**	
b0	0.43	0.031	13.94	0.212	0.003	62.49	
b1	0.04	0.008	4.68	0.037	0.008	4.66	
b2	-0.39	0.132	-2.97	-0.391	0.132	-2.96	
b3	2005.18	0.966	2075.06	2005.177	0.969	2069.57	
access_0_2	-0.0006	0.003	-0.20	-0.0019	0.003	-0.628	
access_2_8	-0.03	0.003	-12.95	-0.03	0.003	-12.73	
access_8_16	-0.10	0.004	-22.91	-0.10	0.004	-22.52	
access_16	-0.17	0.006	-29.62	-0.16	0.006	-28.38	
ln(pop density)	-0.188	0.028	-6.65	0.010	0.004	2.69	ln(density) group2
ln(pop density)^2	0.060	0.009	6.54	-0.016	0.002	-6.64	ln(density) group3
ln(pop density)^3	-0.008	0.001	-6.45	-0.049	0.003	-14.48	ln(density) group4
ln(pop density)^4	0.000	0.000	6.03	-0.074	0.003	-21.72	ln(density) group5
AT	-0.016	0.004	-3.79	-0.015	0.004	-3.60	
BE	-0.064	0.004	-15.11	-0.066	0.004	-15.66	
BG	0.127	0.006	19.99	0.126	0.006	20.06	
CY	0.046	0.012	3.78	0.040	0.012	3.31	
CZ	0.122	0.004	27.35	0.123	0.004	27.39	
DK	-0.048	0.005	-9.03	-0.042	0.005	-7.81	
EE	0.173	0.012	14.14	0.168	0.012	13.93	
EL	0.047	0.006	7.27	0.042	0.006	6.67	
ES	0.054	0.005	10.98	0.052	0.005	10.35	
FI	0.096	0.007	12.88	0.090	0.007	12.04	
FR	-0.038	0.003	-12.03	-0.035	0.003	-11.24	
HR	0.050	0.012	4.23	0.054	0.012	4.52	
HU	0.130	0.007	18.96	0.132	0.007	19.19	
IE	-0.047	0.008	-5.81	-0.043	0.008	-5.35	
IT	-0.006	0.003	-1.92	-0.006	0.003	-1.74	
LT	0.057	0.012	4.93	0.060	0.012	5.25	
LU	-0.108	0.011	-9.81	-0.112	0.011	-10.17	
LV	0.114	0.012	9.35	0.110	0.012	9.07	
MT	0.180	0.012	14.61	0.191	0.012	15.37	
NL	-0.096	0.004	-26.31	-0.095	0.004	-25.45	
PL	0.066	0.004	17.02	0.067	0.004	17.44	
PT	0.090	0.007	12.91	0.084	0.007	12.16	
RO	0.141	0.006	24.54	0.143	0.006	25.08	
SE	0.033	0.006	5.79	0.043	0.006	7.58	
SI	0.063	0.008	7.79	0.066	0.008	8.10	
SK	0.071	0.006	12.03	0.075	0.006	12.50	
UK	-0.057	0.003	-22.06	-0.058	0.003	-22.17	
adjusted R^2	**60.2%**			**59.9%**			

In model 2, however, b_0 has a straightforward interpretation since the “reference region”, with all exogenous variables equal to zero, in this specification is entirely realistic: a German region within the „Blue Banana“ with a low log population density (among the lowest 5% of all NUTS2 regions).

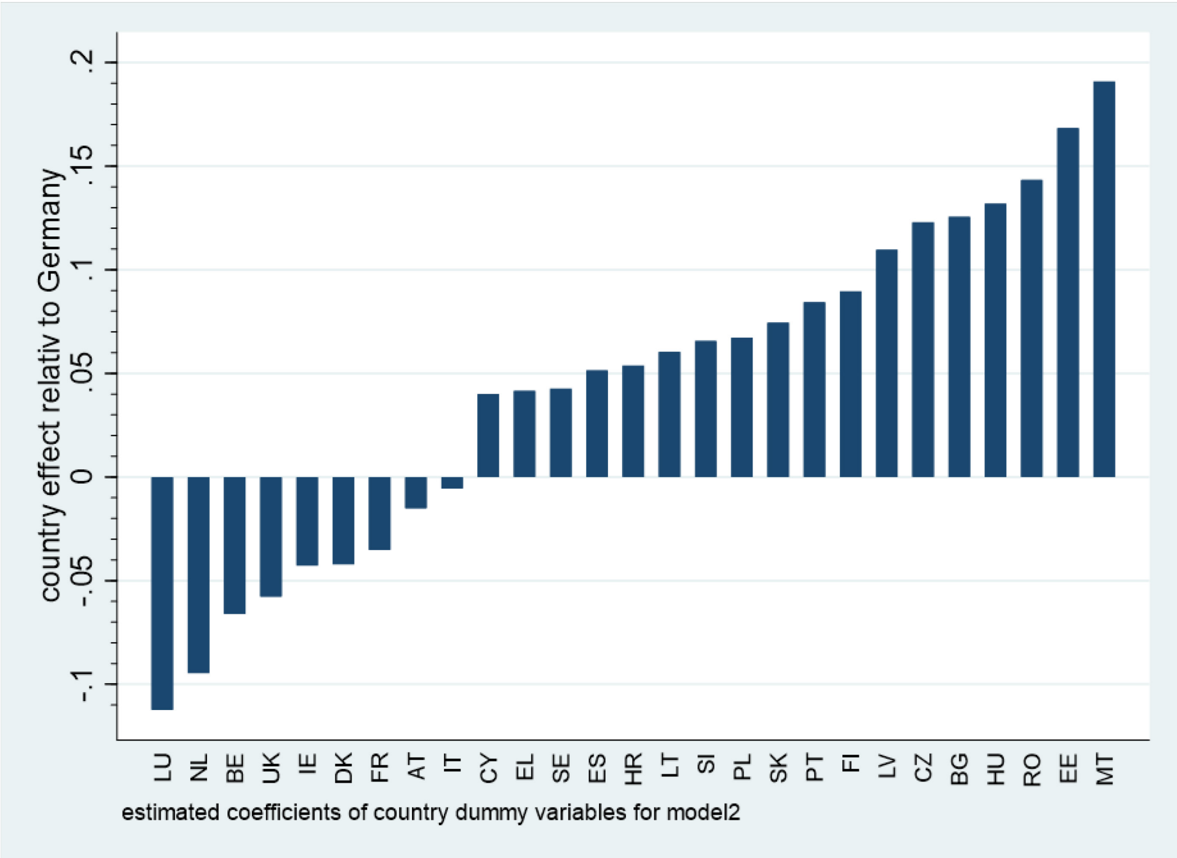
Apart from this difference in the estimated level of b_0 that stems from the different definitions of the population density variables used in the two specifications, they exhibit a very similar estimated logistic time trend as can be seen from Figure 11 below.

Figure 11: Estimated average logistic trend of industrial share among NUTS regions according to models 1 and 2.



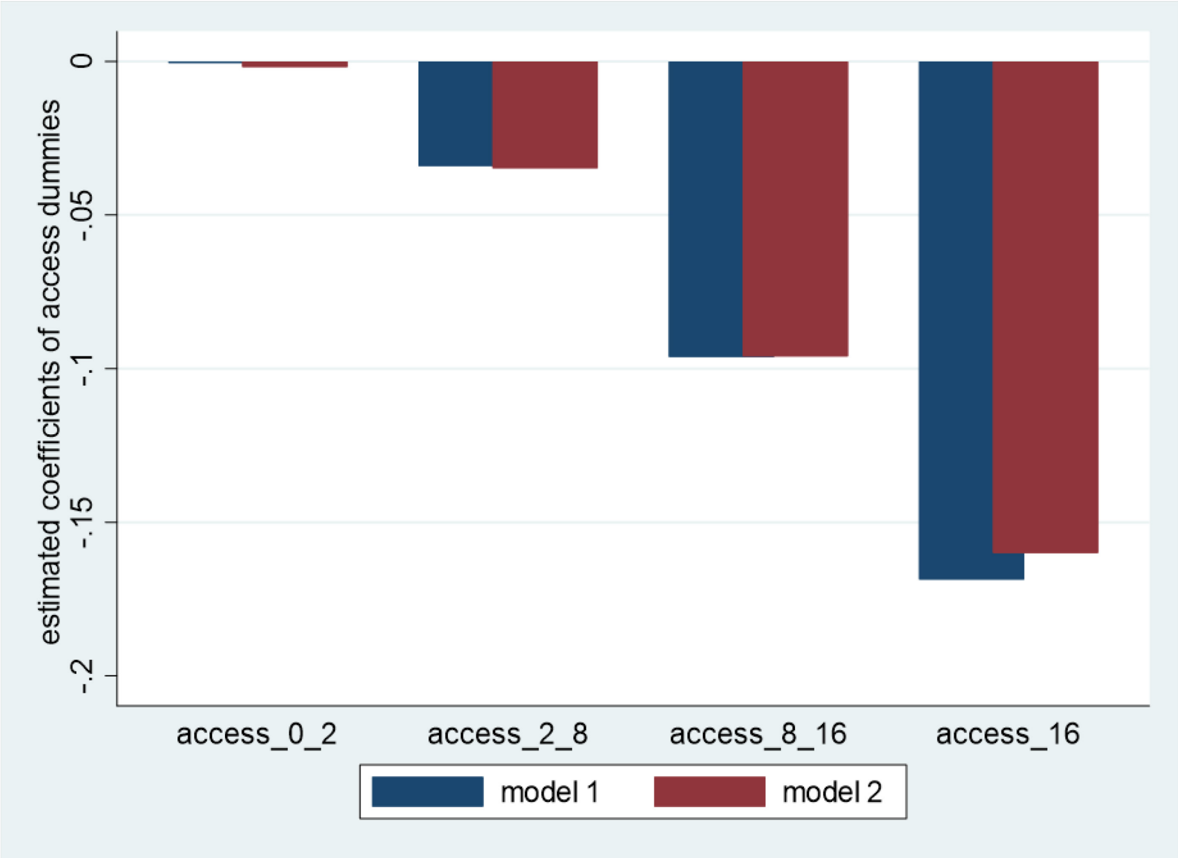
Moreover, both models also imply a similar pattern of estimated country effects. We therefore only visualized the coefficients of the country dummy variables of model 2 as vertical bars in Figure 12. We have ordered the bars in this graph, which has no bar for Germany, as it is the reference country, and is thus omitted in the regression. All estimated coefficients therefore can be interpreted as the average difference in the industrial share of a country's regions relative to the average of regions in Germany. Figure 12 shows that, on average, eastern European regions in Estonia (EE), Romania (RO), Hungary (HU) or Bulgaria (BG) have considerably higher industrial shares than German regions. Germany's Western neighbors, with a similar level of development and production cost, tend to be below the average German level.

Figure 12: Estimated country effects in model 2, relative to the reference country Germany



The estimates for models 1 and 2 also largely agree with respect to the estimated impact of location on a region's expected industrial share. Our model selection procedure favored the measurement of location advantages via "access" dummy variables representing intervals of travel time to the "Blue Banana" core region of the common market. The estimated coefficients of these access dummy variables are shown in Figure 13.

Figure 13: Estimated effect of time-to-Blue-Banana access dummy variables for models 1 and 2.



The reference category here consists of regions within the area of the “Blue Banana”. The bars in Figure 13 thus indicate how regions with successively higher travel times to the nearest large city inside the “Blue Banana” compare to the “insiders”. As can be seen, regions less than two hours away (access_0_2) do not appear to differ from those within the “Blue Banana” However, regions with 2 to 8 hours (access_2_8), 8 to 16 hours (access_8_16) or more than 16 hours (access_16) of travel time exhibit substantial reductions in their average industrial shares, just as trade theory would suggest.

Finally, we turn to the estimated effect of population density in the two models. A high population density is a proxy for the positive agglomerative effects on economic activity generated by metropolitan areas. Models 1 and 2 both employ population density in logarithmic form. Model 2 however does not use log population density directly as a continuous regressor but rather employs dummy variables corresponding to the intervals of the log population density distribution shown above in Figure 6. The implied estimated relationships between the expected industrial share and log population density are contrasted with each other in Figure 14.

Figure 14: Estimated relationship between expected industrial share and log population density.

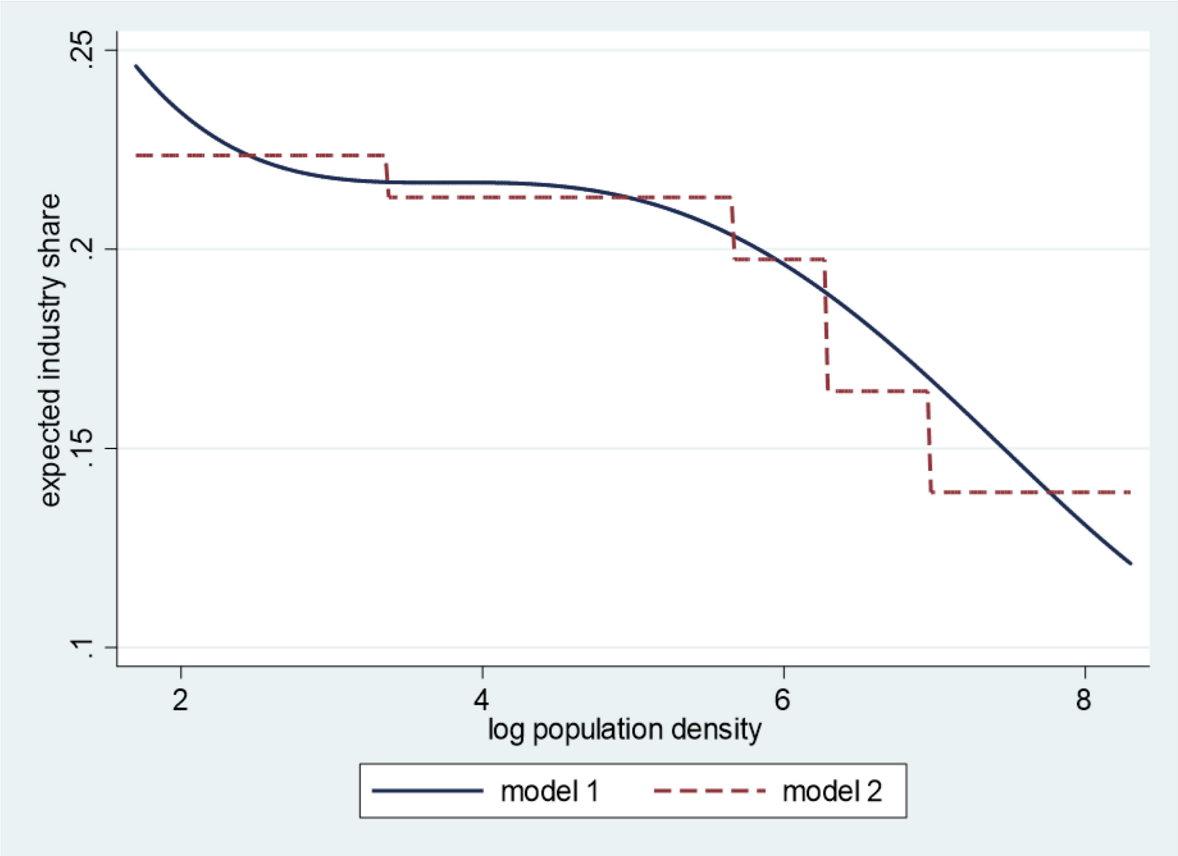


Figure 14 shows that the estimated relationships from both models are quite similar. In model 1 a polynomial of order 4 was used to fit the effect of log population density as a continuous regressor. Model 2, on the other hand, employed a step function approach by estimating the coefficients of dummy variables defined for the indicated intervals of log population density. While the step function, with its “edgy” graph, appears to be the cruder, less flexible approach, the height of its steps (i.e. the coefficient estimates of the dummy variables) are a priori unrestricted and entirely determined by the data. It is thus evidence for the robustness of our findings that the step-function of model 2 and the 4th order polynomial of model 1 paint a very similar picture of the effect of log population density. This estimated effect is basically flat until a level of the log population density of around 5.5, after which the average industrial share tends to noticeably decrease with increasing log population density. Hence, neither of our two alternative specifications shows a positive agglomerative effect. Rather, industrial shares in densely populated regions tend to be lower on average. A possible explanation is that space-consuming industrial activity is facing stiff competition from other uses of scarce land in densely populated metropolitan areas, where all kinds of activities want to enjoy the advantages that large agglomerations have to offer.

7.2.1 Identifying over- and under achieving regions

Given our model estimates, we can calculate for each region the deviation of its expected industry share from its actual industry share in 2015 according to equation (3) above. These residuals can then be used to identify "over achieving" regions with large positive deviations and "under achieving" regions with large negative residuals. The residuals of these regions are shown in Figure 15, based on the estimates for model 2.

Figure 15: Residuals of over-achieving and under achieving regions.

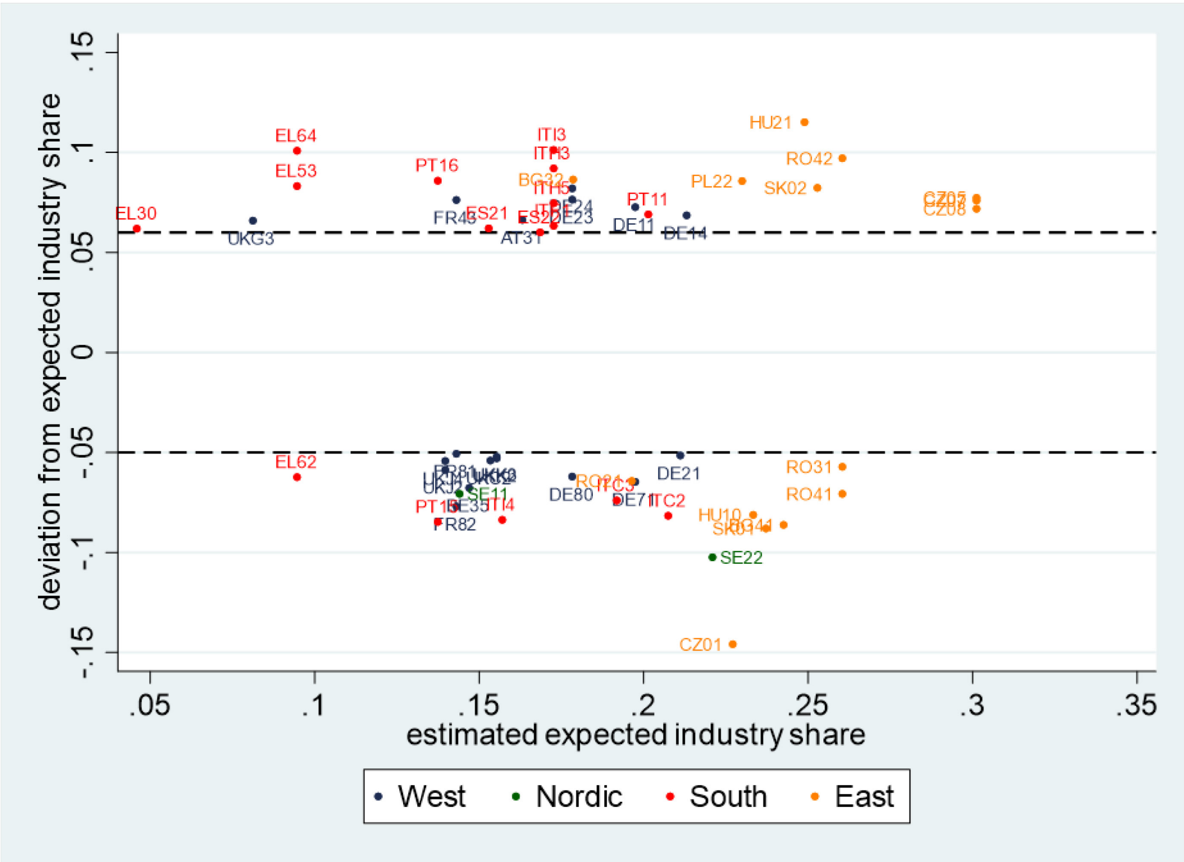


Figure 15 plots these deviations against a region's expected industry share on the horizontal axis. We have plotted only the points of those regions whose residual either exceeds the 90th percentile of the residual distribution (dashed horizontal line at 0.06) or falls below the 10th percentile (dashed horizontal line at -0.05). Hence, only the top 10% and bottom 10% of all regions are shown in the graph. They are also listed in the following Table 6 and Table 7.

Table 6: The bottom 10% of all regions with respect to the deviation of their actual industry share from their expected industry share in 2015 according to model 2

Under Achievers

Region	residual	actual share	expected share
CZ01 – Praha	-0.146	0.081	0.227
SE22 – Sydsverige	-0.102	0.119	0.221
SK01 - Bratislavský kraj	-0.088	0.149	0.237
BG41 – Yugozapaden	-0.086	0.156	0.243
PT15 – Algarve	-0.085	0.053	0.137
ITI4 – Lazio	-0.084	0.073	0.157
ITC2 - Valle d'Aosta/Vallée d'Aoste	-0.082	0.126	0.207
HU10 - Közép-Magyarország	-0.081	0.152	0.233
FR82 - Provence-Alpes-Côte d'Azur	-0.077	0.066	0.143
ITC3 – Liguria	-0.074	0.118	0.192
SE11 – Stockholm	-0.071	0.073	0.144
RO41 - Sud-Vest Oltenia	-0.071	0.190	0.260
BE35 - Prov. Namur	-0.068	0.079	0.147
DE71 – Darmstadt	-0.065	0.133	0.197
RO21 - Nord-Est	-0.064	0.132	0.196
EL62 - Ionia Nisia	-0.062	0.032	0.095
DE80 - Mecklenburg-Vorpommern	-0.062	0.116	0.178
UKJ2 - Surrey, East and West Sussex	-0.059	0.081	0.140
RO31 - Sud – Muntenia	-0.057	0.203	0.260
UKJ4 – Kent	-0.054	0.085	0.140
UKC2 - Northumberland and Tyne and Wear	-0.054	0.099	0.153
UKK2 - Dorset and Somerset	-0.053	0.102	0.155
UKK3 - Cornwall and Isles of Scilly	-0.052	0.103	0.155
DE21 – Oberbayern	-0.052	0.160	0.211
FR81 - Languedoc-Roussillon	-0.051	0.092	0.143

Table 7: The top 10% of all regions with respect to the deviation of their actual industry share from their expected industry share in 2015 according to model2

Over Achievers

Region	residual	actual share	expected share
HU21 - Közép-Dunántúl	0.115	0.364	0.249
ITI3 – Marche	0.101	0.274	0.173
EL64 - Sterea Ellada	0.101	0.196	0.095
RO42 – Vest	0.097	0.358	0.260
ITH3 – Veneto	0.092	0.265	0.173
BG32 - Severen tsentralen	0.086	0.265	0.179
PT16 - Centro (PT)	0.086	0.223	0.137
PL22 – Slaskie	0.086	0.316	0.230
EL53 - Dytiki Makedonia	0.083	0.178	0.095
SK02 - Západné Slovensko	0.082	0.335	0.253
DE24 – Oberfranken	0.082	0.260	0.178
CZ05 – Severovýchod	0.077	0.378	0.301
DE23 – Oberpfalz	0.076	0.255	0.178
FR43 - Franche-Comté	0.076	0.219	0.143
CZ07 - Strední Morava	0.076	0.377	0.301
ITH5 - Emilia-Romagna	0.075	0.247	0.173
DE11 – Stuttgart	0.073	0.270	0.197
CZ08 – Moravskoslezsko	0.072	0.373	0.301
PT11 – Norte	0.069	0.270	0.201
DE14 – Tübingen	0.069	0.282	0.213
AT31 – Oberösterreich	0.066	0.229	0.163
UKG3 - West Midlands	0.066	0.147	0.081
ITF1 – Abruzzo	0.063	0.236	0.173
ES21 - País Vasco	0.062	0.215	0.153
EL30 – Attiki	0.062	0.108	0.046
ES22 - Comunidad Foral de Navarra	0.060	0.229	0.169

It can be seen particularly from Figure 15, which also uses color to indicate Western, Eastern, Southern and Nordic regions, that under- and over-achievers can be found at all levels of the expected industry share and in all geographical zones of the European Union. Indeed, there are several instances were over- and under-achievers belong to the same country and thus face the same national framework conditions. This points to the importance of local conditions and policies for achieving or failing to achieve an industrial share that exceeds expectations. Hence, policy measures aiming at raising industrial shares must take local conditions into account.

This is underscored by a closer inspection of the list of under-achievers in Table 6. Distinctly different types of under-achievers can be identified. There are, for instance, several under-achieving regions that host the respective country’s capital city. Apparently, hosting the main political institutions of a country aggravates the competition for scarce land in such agglomerations and leaves less room for space consuming industrial firms and plants. Other under-achieving regions such as Liguria, Provence-Alpes-Côte d'Azur or Algarve are prime destinations of tourism. Attempts of raising their industry share could build on their well-developed transportation infrastructure but needs to avoid compromising their attractiveness for visitors. Still other regions, such as Yugozapaden or Sud–Muntenia suffer from their remote location both within Europe and their own country. Increasing industrial activity in these regions could build on their cost advantages but would require large scale infrastructure investments to create much improved links to lucrative markets.

8 Conclusions

The aim of our empirical analysis was to estimate the expected industrial share of a region in 2015, given its exogenous conditions, and to confront this expected value with the observed industrial share. Our regression-based approach does not aim at accurate predictions of a region’s industrial share. Instead, it is rooted in theory and conditions on a few important exogenous factors. This allowed us to identify regions which - given their exogenous framework conditions- can be considered as over-achieving and under-achieving with regard to their industrial share. These regions come from all levels of expected industry shares and from all geographical areas of the EU. This suggests that local conditions also appear to considerably matter for a region’s industrial share. Hence, when designing policies for achieving its ambitious industry goals, EU policy makers must make sure that measures do not only improve conditions at-large but also serve the particular needs of the targeted regions. Increasing industrial activity in capitol regions, for instance, must grapple with the fierce local competition for space exacerbated by housing government, parliament, media and lobbyists. The remote under-achieving regions in South-Eastern Europe, on the other hand, can only leverage their cost advantages into more industrial production activity if their transport connection to key markets is sufficiently developed.

References

Abbott, Andrew, and Glauco de Vita. 2013. "Testing for Long-Run Convergence across Regional House Prices in the UK: A Pairwise Approach." Applied Economics 45 (10): 1227–38. http://dx.doi.org/10.1080/00036846.2011.613800 .
Anselin, Luc. 1995. "Local Indicators of Spatial Association—LISA." Geographical Analysis 27 (2): 93–115. http://dx.doi.org/10.1111/j.1538-4632.1995.tb00338.x .
Beenstock, Michael, and Daniel Felsenstein. 2007. "Spatial Vector Autoregressions." Spatial Economic Analysis 2 (2): 167–96. http://dx.doi.org/10.1080/17421770701346689 .
Beenstock, Michael, and Daniel Felsenstein. 2010. "Spatial Error Correction and Cointegration in Nonstationary Panel Data: Regional House Prices in Israel." Journal of Geographical Systems 12 (2): 189–206. http://dx.doi.org/10.1007/s10109-010-0114-8 .
Carmona, Juan, and Joan R. Rosés. 2012. "Land Markets and Agrarian Backwardness (Spain, 1904-1934)." European Review of Economic History 16 (1): 74–96. http://dx.doi.org/10.1093/ereh/her001 .
Deininger, Klaus, and Derek Byerlee. 2011. "Rising Global Interest in Farmland. Can it Yield to Sustainable and Equitable Benefits?" Washington: The World Bank.
European Commission. 2016. "Agricultural Land Prices and Rents data for the European Union." http://ec.europa.eu/eurostat/web/agriculture/methodology .
Falk, Barry. 1991. "Formally Testing the Present Value Model of Farmland Prices." American Journal of Agricultural Economics 73 (1): 1–10. http://dx.doi.org/10.2307/1242877 .
Federal Statistical Office of Germany. 2017. "Land- und Forstwirtschaft, Fischerei: Kaufwerte für landwirtschaftliche Grundstücke 2016". https://www.destatis.de/DE/Publikationen/Thematisch/Preise/Baupreise/KaufwerteLandwirtschaftlicheGrundstuecke.html